DOCUMENT RESUME

ED 137 363 TM 006 164

AUTHOR Rudner, Lawrence M.
TITLE A Closer Look at Latent Trait Parameter Invariance.
INSTITUTION Model Secondary School for the Deaf, Washington, D.C.
PUB DATE [May 77]
NOTE 17p.; Paper presented at the Annual Meeting of the New England Educational Research Organization (Manchester, New Hampshire, May 4-7, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Ability; *Cognitive Measurement; Error Patterns; *Mathematical Models; *Measurement; Probability; Response Style (Tests); Testing
IDENTIFIERS Item Characteristic Curve Theory; *Latent Trait Models; *Latent Trait Theory

ABSTRACT

Using conventional mental test theory, item parameters of an aptitude or achievement test vary with each group of examinees, and as such are somewhat limited in their use and interpretation. Within the last 25 years, measurement models have emerged in which item parameters are considered to be invariant. Generically referred to as latent trait mental measurement or item characteristic curve theory, these models have precipitated the development of sophisticated tailored testing techniques which capitalize on the item parameter invariance feature. This paper examines the invariance feature for the Birnbaum 2 and 3 parameter logistic models. The investigation is in two parts, the first theoretical, the second empirical. Even with somewhat inadequate data, the empirical investigation supported the theoretical considerations. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

A CLOSER LOOK AT

LATENT TRAIT PARAMETER INVARIANCE

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Lawrence M. Rudner

The Model Secondary School for the Deaf
Gallaudet College, Washington, D. C.

Paper presented at the Annual Meeting of the
New England Educational Research Organization
Manchester, New Hampshire, May 4 - 7, 1977

TM006 164

A Closer Look at Latent Trait Parameter Invariance

Using conventional mental test theory, item parameters of an aptitude or achievement test vary with each group of examinees, and as such are somewhat limited in their use and interpretation. Within the past 25 years, measurement models have emerged in which item parameters are considered to be invariant (Lord, 1952; Rasch, 1960; Birnbaum, 1968; Urry, 1970). Generically referred to as "latent trait mental measurement" or "item characteristic curve theory," these models have precipitated the development of sophisticated tailored testing techniques which capitalize on the item parameter invariance feature.

In this paper, I will examine this invariance feature for the Birnbaum 2 and 3 parameter logistic models (Birnbaum, 1968). The investigation will be in two parts, the first theoretical, the second empirical.

A Theoretical Examination of Parameter Invariance

Both models under consideration have been developed in the literature (Jensema, 1972; Lord and Novick, 1968), and will not be fully described here. The item characteristic curve (ICC) in Figure 1 shows, for Birnbaum's 3 parameter logistic model, the regression of ability θ, against the probability of a correct response,

$$P(\theta) = c_g + (1-c_g)[1 + \exp(-1.7a_g(\theta-b_g))]^{-1} \qquad (1)$$

The author is indebted to Carl Jensema, Leo Min, and John Convey for their assistance with the first draft of this report, to the Office of Demographic studies, Gallaudet College, Washington, D.C., for supplying the data used in the study, and to Jacqueline Cox and Joyce Robinson for typing the manuscript.

for a particular item g; where

b_g, the item difficulty, is the point of inflection indicating the location of the curve along the θ axis;

a_g, the item discrimination, is directly related to the slope of the curve at b_g;

c_g, the probability of guessing the correct response, is the lower asymptote of the curve.

In the Birnbaum 2 parameter logistic model, c_g=0, for all items, and equation 1 becomes

$$P(\theta) = [1 + \exp(-1.7a_g(\theta - b_g))]^{-1} \quad (2)$$

The parameters b_g, a_g, and c_g are considered invariant in that the shape of any given ICC (for a particular item) is theoretically invariant across populations and across tests. That is, if an item is transplanted into a different test of the same ability or administered to a different group of examinees, the shape of the ICC will not change.

The numerical values for a_g and b_g, however, are dependent upon the scale used to define particular values of θ. For lack of a natural metric, these values of θ are usually represented as deviations from the mean true ability in a pre-tested population. Thus, if item parameters are separately calibrated on two groups of examinees whose ability distributions differ, different numerical values to represent each a_g and b_g parameter will be obtained. This situation is illustrated in Figure 2.

The two rigid θ scales may be represented by θ' and θ''. Let a_g', b_g', and c_g' represent the numeric values obtained using the θ' scale, and let a_g'', b_g'', and c_g'' represent the values using the θ'' scale. Clearly, the

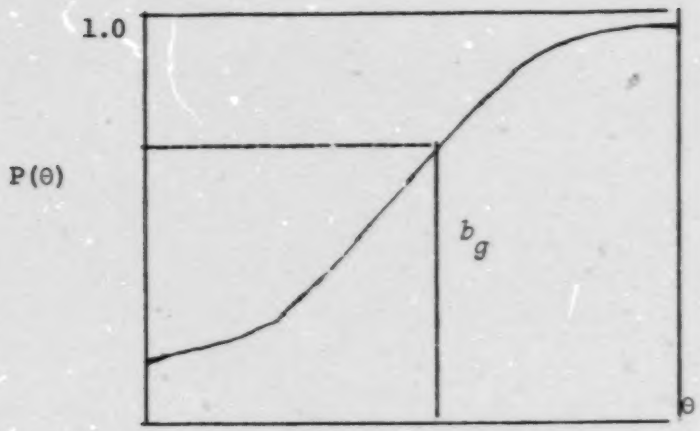

Figure 1. A hypothetical item characteristic curve.

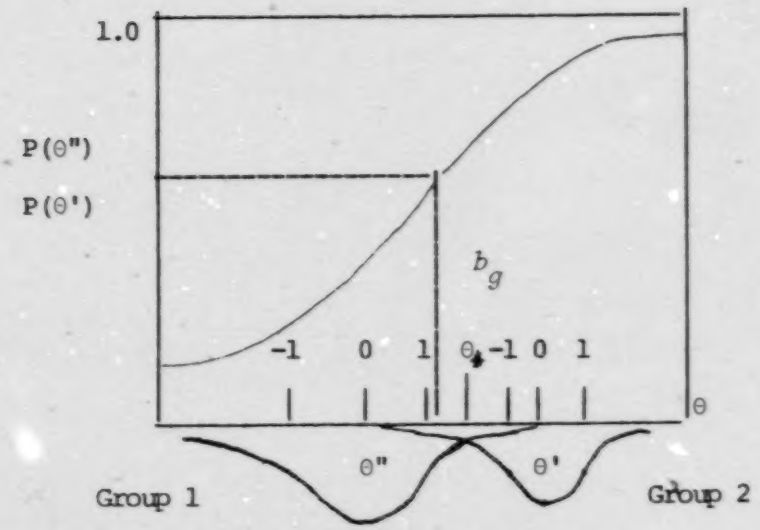

Figure 2. A hypothetical item characteristic curve with two ability scales.

$b_g' \neq b_g''$ and $a_g' \neq a_g''$. In theory, $c_g' = c_g''$; however, because of the difficulty of obtaining good c_g parameter estimates, this cannot be expected in practice.

Consider an examinee whose true ability is θ_* and who is a member of both groups. If the shape of the ICC is indeed invariant, this examinee will have the same probability of responding correctly regardless of the θ scale. That is,

$$P(\theta_*') = P(\theta_*'') \tag{3}$$

where θ_*' and θ_*'' are the numerics used to describe θ_* on the θ' and θ'' scales respectively.

New numerical values for a_g and b_g should be predictable with each change in the θ scale. I will derive and discuss such prediction equations for both the Birnbaum 2 parameter and the Birnbaum 3 parameter logistic models, and finally will describe an empirical investigation.

Birnbaum's 2 Parameter Model

By definition, (equation 2)

$$P(\theta_*'') = [1 + \exp(-1.7a_g''(\theta_*'' - b_g''))]^{-1}$$

and

$$P(\theta_*') = [1 + \exp(-1.7a_g'(\theta_*' - b_g'))]^{-1}$$

Considering θ'' and θ' not equal to b_g'' and b_g'; substituting into (3) and inverting, we obtain:

$$1 + \exp(-1.7a_g''(\theta_*'' - b_g'')) = 1 + \exp(-1.7a_g'(\theta_*'' - b_g''))$$

Subtracting 1 and taking the natural logarithms,

$$-1.7a_g''(\theta_*'' - b_g'') = -1.7a_g'(\theta_*' - b_g')$$

Then

$$\hat{a}_g'' = \left[\frac{\theta_*' - b_g'}{\theta_*'' - b_g''}\right] a_g' \tag{4}$$

and

$$\hat{b}_g'' = \left[\frac{a_g'}{a_g''}\right] b_g' + \frac{a_g''\theta_\star'' - a_g'\theta_\star'}{a_g''} \tag{5}$$

Carets ($\wedge$) are used to indicate that these are predicted values. Note that both (4) and (5) are linear relationships. The same types of linear relationships hold for Lord's 2 parameter normal ogive model (see Lord and Novick, 1968, Chapter 16.11).

Birnbaum's 3 Parameter Model

In theory, the values for a_g and b_g do not change when a guessing parameter, c_g, is added. As such, equations 4 and 5 should hold for the three parameter logistic model as well. However, the question will be answered allowing for $c_g' \neq c_g''$.

First $\theta_\star'$ and $\theta_\star''$ are selected such that $P(\theta_\star')=P(\theta_\star'')=K$, $0<K<1$. Thus,

$$\theta_\star' = b_g' + \ln\left[\frac{K - c_g'}{1 - K}\right](1.7a_g')^{-1} \tag{6}$$

and

$$\theta_\star'' = b_g'' + \ln\left[\frac{K - c_g''}{1 - K}\right](1.7a_g'')^{-1} \tag{7}$$

Let K' and K" denote $\ln\left[\frac{K - c_g'}{1 - K}\right]$ and $\ln\left[\frac{K - c_g''}{1 - K}\right]$ respectively. Then solving (6) and (7) for 1.7,

$$\frac{K''}{(\theta'' - b_g'')a_g''} = \frac{K'}{(\theta' - b_g')a_g'} = 1.7$$

Then

$$\hat{a}_g'' = \left[\frac{\theta_\star' - b_g'}{\theta_\star'' - b_g''}\right]\left[\frac{K''}{K'}\right] a_g' \tag{8}$$

and

$$\hat{b}_g'' = \left[\frac{K''\, a_g'}{K'\, a_g''}\right] b_g' + \frac{\theta_\star'' a_g'' K' - \theta_\star' a_g' K''}{a_g'' K'} \qquad (9)$$

Again, these are linear functions. Note that when $c_g' = c_g''$, $K' = K''$ and equations 8 and 9 reduce to equations 4 and 5 respectively.

Discussion

This demonstration has shown that with both the Birnbaum 2 and 3 parameter models, (1) a_g'' is a linear function of a_g' with intercept equal to zero and (2) b_g'' is a linear function of b_g' with intercept not necessarily equal to zero.

In equations 4 and 8, one may note that

$$\frac{\theta_\star' - b_g}{\theta_\star'' - b_g} \qquad (10)$$

is a function solely of two metrics, θ' and θ'', used to describe $\theta_\star'$ and b_g', and $\theta_\star''$ and b_g''. As such, (10) and consequently (4) and (5) are constant for all items within the test and (8) and (9) will vary slightly when between group guessing parameters are unequal.

With the Rasch model, $c_g' = c_g'' = 0$ and $a_g' = a_g''$ for all items, and (10) is again constant. As with the Birnbaum model, the difficulty (easiness) parameter of the Rasch model can also be expected to be numerically invariant over populations. This is precisely what Anderson, et al. (1968) and Tinsley and Dawis (1972) found in empirical investigations.

I have shown that the numerical values for a_g and b_g of both Birnbaum models change linearily as a function of the change in metric. The parameters are invariant in that theoretically one can always recalibrate the a_g and b_g numeric values to a different θ metric without any loss in interpretation.

An Empirical Investigation of Parameter Invariance

The parameter invariance has been considered theoretically in the preceding pages. In this part of the paper, it will be treated empirically. Specifically, I will show that a change in the θ metric will precipitate linear changes in the a_g and b_g values.

Procedure

Item responses to the Level II Spelling Subtest of the Stanford Achievement Test, Hearing-Impaired Version (SAT-HI) by 2,609 students were made available through the courtesy of the Office of Demographic Studies, Gallaudet College, Washington, D.C. Each of the 47 items within this subtest provides one correct and three alternative responses.

The 2,609 cases were a part of a stratified random sample of 6,872 students in the United States' special education programs for the hearing-impaired taking the entire SAT-HI battery in the spring of 1974. Stratification was conducted on type of program and enrollment size of the institution.

The examinees were divided into two groups based on chronological age on the day of testing. The younger group, group L, contained 1,415 subjects; the older group, group H, 1,194.

Parameter estimates were calibrated on each group separately using an iterative minimum chi square estimation procedure developed by Urry (1975). As suggested by Schmidt and Gugel (1975), a minimum of 1,000 subjects and 60 items are needed to obtain satisfactory parameter estimates with the Urry procedure. Since the number of items available on the SAT-HI is less than 60, estimates from both groups may contain somewhat more random error than recommended.

The data were analyzed to test the null hypotheses that (1) with the intercept set equal to zero, there is no strong linear regression of a_g^L on a_g^H, and (2) there is no strong linear regression of b_g^L on b_g^H. Fairly restrictive R^2s>.75 were sought. To provide additional information, these regressions were plotted (see Figures 3 and 4) and the product-moment correlations of observed and predicted parameter values were computed.

Results

In the process of parameterization, 16 and 2 subjects were dropped from groups L and H respectively, because θ estimates were unattainable. Item parameter estimates were obtained on a total of 41 items common to both groups. These estimates and the $\hat{a}_g^H$ and the $\hat{b}_g^H$ values are shown in Table 1. The bivariate product-moment correlations between $\hat{a}_g^H$ and a_g^H and between $\hat{b}_g^H$ and b_g^H are .677 and .929, respectively.

Table 1

Observed and Predicted a_g and b_g Values

on a 41-Item Test

PREDICTED		GROUP H N=1,194			GROUP L N=1,415		
$\hat{a}_g$	$\hat{b}_g$	a_g^H	b_g^H	c_g^H	a_g^L	b_g^L	c_g^L
1.57	0.18	1.70	-0.02	0.21	1.49	0.91	0.24
2.17	-1.27	2.03	-1.24	0.04	2.06	-0.74	0.07
1.43	-0.68	1.68	-0.73	0.10	1.36	-0.07	0.11
1.33	-1.16	1.24	-0.92	0.06	1.26	-0.62	0.10
1.69	-1.07	1.73	-1.08	0.05	1.60	-0.52	0.24
1.26	-0.54	1.46	-1.08	0.06	1.20	0.09	0.23
1.72	-1.12	2.13	-0.98	0.02	1.63	-0.57	0.09
1.36	-0.56	1.54	-0.71	0.05	1.29	0.06	0.17
1.30	0.32	1.28	-0.01	0.16	1.23	1.06	0.22
1.65	-0.11	1.32	-0.08	0.14	1.57	0.58	0.24
0.85	-0.13	1.02	-0.51	0.13	0.81	0.55	0.27
1.17	-0.37	0.89	-0.49	0.11	1.11	0.28	0.29
0.77	-0.16	0.71	0.27	0.16	0.73	0.52	0.12
1.17	-0.92	1.42	-0.84	0.05	1.11	-0.35	0.09
0.76	-0.79	0.95	-0.63	0.05	0.72	-0.20	0.07
1.22	-0.06	1.08	-0.34	0.10	1.16	0.63	0.21
1.19	0.09	1.40	-0.33	0.16	1.13	0.80	0.26
1.26	-0.27	1.25	-0.35	0.22	1.20	0.39	0.22
1.06	-0.16	1.05	-0.22	0.05	1.01	0.52	0.14
1.36	-0.92	1.48	-0.89	0.06	1.29	-0.35	0.09
2.00	0.37	1.77	0.28	0.16	1.90	1.12	0.17
1.88	0.07	1.15	0.06	0.15	1.78	0.78	0.23
1.02	-0.96	0.77	-0.84	0.05	0.97	-0.39	0.09
1.12	-0.48	1.42	-0.58	0.10	1.06	0.16	0.14
1.72	-1.00	1.30	-0.85	0.06	1.63	-0.44	0.08
1.13	-0.57	1.16	-0.45	0.05	1.07	0.05	0.08
1.22	-0.69	1.21	-0.65	0.05	1.16	-0.08	0.10
1.60	-0.51	1.72	-0.47	0.10	1.52	0.12	0.17
1.23	0.52	1.20	0.33	0.18	1.17	1.29	0.26
2.27	1.19	1.63	1.72	0.41	2.15	2.05	0.32
1.41	-0.72	1.27	-0.82	0.06	1.34	-0.12	0.12
1.25	-0.70	1.20	-0.79	0.06	1.19	-0.10	0.08
1.17	0.02	1.89	0.26	0.14	1.11	0.72	0.20
1.42	-0.23	1.93	-0.25	0.09	1.35	0.44	0.13
0.70	0.40	0.75	0.41	0.12	0.66	1.15	0.16
1.39	-0.29	1.67	-0.09	0.19	1.32	0.37	0.14
1.60	1.10	2.17	1.18	0.26	1.52	1.95	0.29
1.39	-0.16	1.51	-0.08	0.22	1.32	0.52	0.23
0.99	0.46	1.12	0.06	0.18	0.94	1.22	0.26
1.83	0.12	1.56	0.54	0.23	1.74	0.84	0.21
1.21	0.45	1.28	0.88	0.26	1.15	1.21	0.24

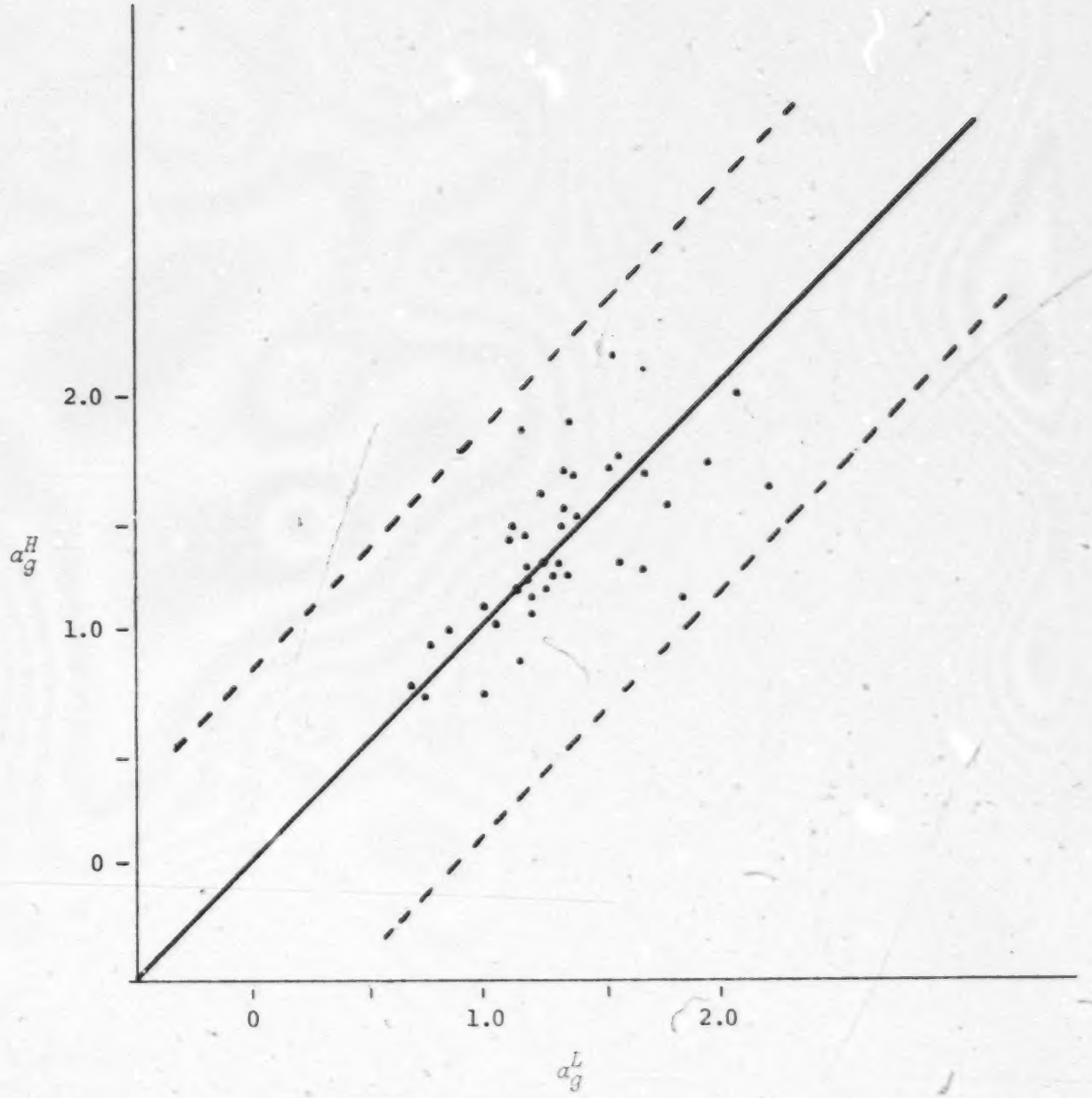

Figure 3. Bivariate frequency plots of item discrimination values a_g from high and low groups.

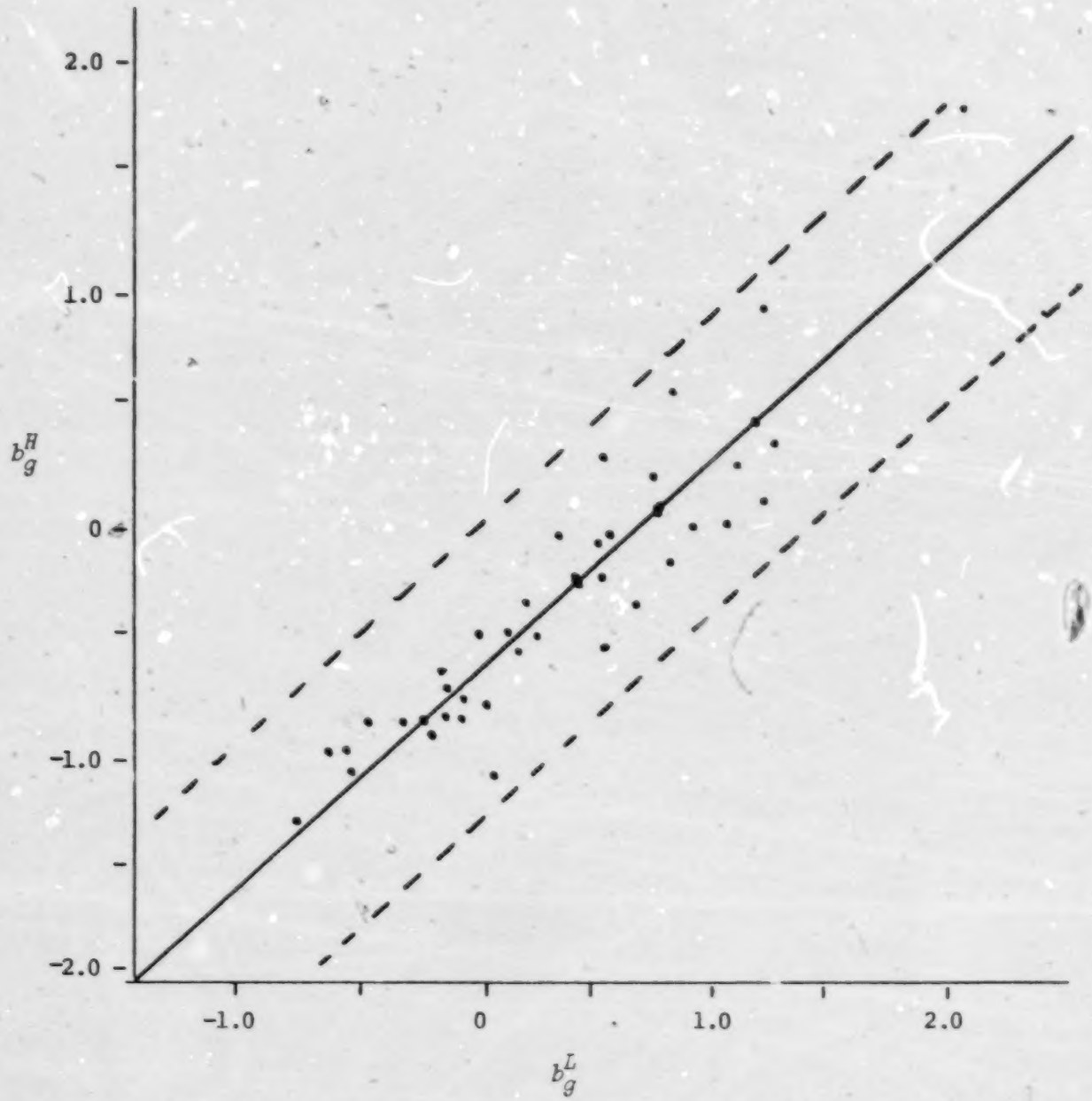

Figure 4. Bivariate frequency plots of item difficulty values b_g for high and low groups.

Table 2 shows the summary tables for the a_g and b_g regressions. It should be noted that the a_g and b_g regressions accounted for large percentages of the parameter variance (96% and 86% respectively). Thus, one can reject the hypotheses that there is no strong linear relationship between a_g values with the intercept set equal to zero and there is no linear relationship between b_g values at $R^2 > .75$.

Table 2

Regression Summary Tables

Regression of a_g^H on a_g^L (fixed intercept)		Regression of b_g^H on b_g^L	
Multiple R	.97976	Multiple R	.92919
R Squared	.95993	R Squared	.86339
Standard Error of Estimation	.29121	Standard Error of Estimation	.23883
Slope (Unstandardized)	1.05408	Slope	.87973
Intercept	.00000	Intercept	-.61600

The regressions of a_g^L on a_g^H and of b_g^L on b_g^H are illustrated in Figures 3 and 4. The solid lines represent the predicted values of $\hat{a}_g^H$ and b_g^H and the dotted lines represent the 95% confidence interval (±1.96 times S.E. est.).

Discussion

Although the parameter estimates may contain more than the recommended amounts of error variance, the presence of strong linear relationships has been shown between a_g values and between b_g values when calibrated on different θ scales. In effect, this demonstrates evidence for the invariance of the discrimination and difficulty latent trait parameters, even with fewer items than recommended.

An observed value deviating from a predicted value indicates one or more of the following:

1. a violation of the unidimensionality restriction,
2. a violation of the local independence restriction,
3. a poor parameter estimate when calibrated on one or both groups.

In light of the test size, the non-perfect correlations and deviations from the prediction lines are perhaps attributable to the third reason above. With existing parameter estimation programs, the a_g estimates constantly contain more random error than the b_g estimates. In addition, there is some restriction in range with regard to this parameter. Hence, the lower correlation for the a_g parameters is to be expected.

Conclusion

In this paper, I have developed the parameter invariance concept for the Birnbaum 2 and 3-parameter logistic models. On a theoretical level, I showed that the numeric used to describe a_g and b_g is dependent upon the θ scale definition, and that through linear equations, new values of a_g and b_g can be predicted when the θ scale is redefined.

Even with somewhat inadequate data, the empirical investigation supported the theoretical considerations. At conservative values, the study demonstrated that a change in the θ metric precipitated linear changes in the values and in the b_g values. The predicted values of $\hat{a}_g^H$ and $\hat{b}_g^H$ correlated with, and did not overly deviate from, the observed values of a_g^H and b_g^H, and large percentages of the parameter variances were explained by the linear equations.

The presence of these linear functions supports the validity of the model for the test and for the examined samples. That is, assuming that satisfactory parameter estimates can be obtained, this procedure can be used to check for violations of the unidimensionality and local independence latent trait theory restrictions. The presence of near perfect linear relationships between samples would indicate that the restrictions have been met. Its absence between samples would suggest either a restriction violation or poor parameter estimates.

REFERENCES

Anderson, J., Kearney, G.E., & Evertt, A.U. An Evaluation of Rasch's Structural Model for Test Items. British Journal of Mathematics and Statistical Psychology, 1968, 21, 231-238.

Birnbaum, A. Some Latent Trait Models and their Use in Inferring an Examinees Abilities. In F.M. Lord and M.R. Novich, Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Jensema, C.J. An Application of Latent Trait Mental Test Theory to the Washington Pre-College Testing Battery. Unpublished doctoral dissertation, University of Washington, 1972.

Lord, F.M. A Theory of Test Scores. Pyschometric Monograph, 1952, No. 7.

Lord F.M. & Novick, M.R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Paedoggogishe Institute, 1960.

Schmidt, F.L. & Gugel, J.F. The Urry Item Parameter Estimation Technique: How Effective? Paper presented at the symposium: Computers and Testing: Steps Toward the Inevitable Conquest, at the American Psychological Association Convention, Chicago, August 31, 1975.

Tinsley, H.E. & Dawis, R.V. An Investigation of the Rasch Simple Logistic Model: Sample Free Item and Test Calibration. Educational and Psychological Measurement, 1975, 35, 325-339.

Urry, V.W. A Monte Carlo Investigation of Logistic Mental Test Models. Unpublished doctoral dissertation, Purdue University, 1970.

Urry, V.W. Ancillary Estimators for the Item Parameters of Mental Test Models. Paper presented at the symposium: Computers and Testing: Steps Toward the Inevitable Conquest, at the American Psychological Association Convention, Chicago, August 31, 1975.